R E P O R T R E S U M E S

ED 012 920 AL 000 633

STRESS, ACCENT, AND EMPHASIS.

BY- VANDERSLICE, RALPH

PUB DATE 28 JUL 67

EDRS PRICE MF-$0.25 HC-$0.68 17P.

DESCRIPTORS- *COMPUTATIONAL LINGUISTICS, INTONATION, *SYLLABLES, *SEMANTICS, SYNTAX, ENGLISH, PHONEMES, DIACRITICAL MARKING, *PRONUNCIATION, *READING, READING PROCESSES, GRAPHEMES, SPEECH, SPEECH SKILLS, READING MACHINES,

IN ORDER TO MAKE A MACHINE THAT CONVERTS WRITTEN TEXT INTO SPOKEN LANGUAGE (READING MACHINE), IT WOULD BE NECESSARY TO MAP SENTENCES OF WRITTEN ENGLISH ONTO CORRESPONDING SENTENCES OF SPOKEN ENGLISH, IN A CONVERSION WHICH THE AUTHOR CALLS "SYNTHETIC ELOCUTION." IN THIS TYPE OF CONVERSION, THE ASSIGNMENT OF PROSODIC FEATURES TO SENTENCES IN CONNECTED TEXTS REQUIRES THE APPLICATION OF RULES WHICH ARE OPERABLE ACROSS SENTENCE BOUNDARIES. THEORETICALLY, A MODEL READING MACHINE MIGHT HAVE A CHARACTER RECOGNIZER AS THE INPUT AND AND A SPEECH SYNTHESIZER AS THE OUTPUT AND WOULD INCORPORATE PROGRAMS FOR SYNTACTIC AND SEMANTIC ANALYSIS, WORD PRONUNCIATION, AND PROSODIC FEATURE ASSIGNMENT. MECHANICAL SPEECH SYNTHESIS BY RULES IS COMPLICATED BY A TWO-PART PROBLEM--(1) DETERMINING THE MINIMAL SET OF PROSODIC FEATURES WHICH HAVE TO BE ASSIGNED, AND (2) DETERMINING THE FORM OF THE RULES WHICH WILL CORRECTLY ASSIGN THEM. THE AUTHOR FOCUSES ON SOME ASPECTS OF THE PROSODIC MODEL AND PROPOSES A MORE THOROUGH MEANS OF DIFFERENTIATING SUCH HOMOPHONOUS EXPRESSIONS AS "LIGHTHOUSE KEEPER" AND "LIGHT HOUSEKEEPER." IT SEEMS THAT ACCENT PLACEMENT IN ENGLISH CANNOT ALWAYS BE DETERMINED WITHIN THE LIMITS OF THE SENTENCE. RATHER IT IS PARTIALLY SUSCEPTIBLE TO PREDICTION BY "CONTEXT SENSITIVE RULES." THIS PAPER WAS READ AT THE 1967 SUMMER MEETING OF THE LINGUISTIC SOCIETY OF AMERICA. (FB)

Linguistic Society of America
Summer Meeting

Ann Arbor, Michigan
28 July 1967

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

Stress, Accent, and Emphasis

Ralph Vanderslice

UCLA

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY Ralph Vanderslice

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

ED012920

An adequate prosodic model of English is crucial for orthographic-to-phonetic conversion of the sort I've called "synthetic elocution"—which seems a convenient term to designate a form of speech synthesis by rule where the input comprises orthographic texts and the output is PROSODICALLY as well as orthoepically correct.

The function which maps sentences of a written English text onto corresponding sentences of spoken English is of considerable interest. It has practical interest for such applications as a high performance reading machine for the blind, for example; and it has theoretical interest as a model of some crucial aspects of linguistic competence. Correct assignment of prosodic features to sentences in connected texts requires, far more often than not, the application of rules which operate across sentence boundaries—as I hope to show.

A conceptual model of a reading machine might have a character recognizer as the input and a speech synthesizer as the output. Between these, in a more or less serial order, would have to be programs for syntactic and semantic analysis, word pronunciation, and prosodic feature assignment. (See figure 1.)

It's convenient to start from the point where some of these tasks have been done—such as character recognition, parsing, semantic analysis (esp. anaphora and discourse equivalence), and grapheme-to-phoneme conversion—so that we start with an input comprising a concatenation of

AL 000 633

words with pronunciations (including word stress) assigned, and with discrete signals for punctuation marks, italics, new paragraphs, etc; and, along with this text, categorizations indicating the syntactic and semantic relations. So far as I know, none of these steps can be done really well at present by automatic procedures (although I have been able to try out some prosodic rules by hand simulation upon the output of the Kuno grammar as modified by John Olney's very sophisticated post-processing programs, and I'm extremely impressed with what CAN be done). But the point here is not how to build a real reading machine, but rather we are talking in terms of IF ideally efficacious procedures were available for these tasks, how would you get from there to the point where conventional speech synthesis by rule takes over--where the synthesizer requires an input of phoneme strings with prosodic markers attached. This is really two questions: (1) what is the minimal set of prosodic features which have to be assigned, and (2) what is the form of the rules which will correctly assign them? In practice these become interdependent. I'd like to present today some aspects of a prosodic model--in particular the accentual features, as opposed to the terminal pitch contours or the pausal or the tessitural features. This prosodic model is extremely eclectic, and my debt to Pike; Trager and Smith; Chomsky, et alia; Kurath; Halliday; and particularly to Bolinger and Hockett will be obvious.

Approaching the prosodic model from the view point of orthographic-to-phonetic conversion affords a non-arbitrary basis for a "cut-off" point between a minimal set of features and a full range of prosodic and paralinguistic signals and cues. If we restrict the model at first to prosodic features needed for correct reading of "straight factual prose" it will be considerably simpler than what is needed for fictive, dramatic, and poetic texts. I would argue that this distinction is coterminous

with that between the grammatical and the attitudinal and emotional uses of prosodic cues, but there are many who take a narrower view of grammar. For instance there is a widespread belief (exemplified in Chomsky, Halle, and Lukoff) that contrastive accent can be left out of descriptions of English prosody. In fact, contrast and synonomy relations, along with topic-comment theme-rheme relations, form the crux of the prosodic features assignment problem.

Before going into that, however, let us start by considering that old cliché, the compound noun. The first example is from Jonson (1640),

(1)(i) tènnis còurt kèeper

the other of more recent vintage

(ii) shóe-string-tìp machìne òperàtor.

I use these markings because of their mnemonic value and because it is convenient to present my model as a reinterpretation of the so-called "stresses" of the T-S system. Note however on operator that there is no distinction between the "degrees" of stress shown on the first and third syllables--both are marked by the grave sign meaning here "stressed, but unaccented". This is of course tantamount to a claim that the "secondary stress" of the T-S system does not in fact occur after nuclear accent and in the same sense group, a claim which I think is quite in line with the data. Of course this follows also by definition from designation of the rightmost accent in a sense group as the nuclear accent--a definition which permits us to drop some other features of the T-S system (such as the restriction on the number of significant occurrences of pitch within a sense group).

To say that no other syllables in a sense group after the nuclear-accented one can be accented is, I realize, an iconoclastic assertion. To take up just a few of the most sacrosanct idols, it asserts that 'black,bird as a compound noun and 'black 'bird under contrastive accent shift, as in "that's a 'blǎck 'bìrd|not a rěd ŏne" are homophonous. The same claim holds for white house, black board, English teacher, grand father figure, etc. The whole idea of these supposed stress distinctions, and with it the notion of plus juncture, invite, as Sledd said, the use of Occam's razor.

With respect to that old bore who tends the lighthouse and his frail friend the housekeeper--an example still being picked over (by Lieberman for example) although Sledd complained it was "tiresome" in 1956--it is no doubt true as Bolinger and Gerstman showed that sufficient disjuncture can resolve the ambiguity, but that fact has no more to do with establishing the minimal prosodic features for American English than the fact that clarity-standard articulation can distinguish ladder from latter has to do with the description of the segmental level. All the evidence points to the fact that BOTH cases are homophonous in normal conversation. So then at the input of the accent-assignment rules, each word is recorded with appropriate stress marks as

(2) 'keeper, ma'chine, 'ope,rator

where the "half"-stress mark on the penult of operator indicates a syllable that is stressed in terms of duration, etc., but which is not in the normal course of things eligible for pitch accent assignment. The class of words like this with falling stress is a very large and familiar one: as

(3) ˈverti,go, ˈcan(,)not, ˈsolip,sism, ˈobfus,cate, ˈfructi,fy, ˈmacra,me, etc.

If you are addicted to the use of digits over the syllables (that wonderful system where more is less) I assure you that they can be transliterated by a fairly simple algorithm — a sort of stress readjustment rule — which shucks off the excess structure and maps X degrees of stress onto the desired two binary distinctions which I postulate as incorporated in the IPA full and half stress marks: presence of EITHER stress ~~w~~ mark indicating those syllables normally receiving the greater duration which I am here calling stress, and FULL-stress as opposed to halfstress marks indicating those STRESSED syllables having, as Bolinger puts it, potential for pitch accent. The distribution of these stresses in the sentence then is lexically determined and is altered only under contrastive accent shift. This captures the empirically well-attested fact that most of what people have been discussing in terms of "stress" is really a matter of pitch obtrusions, (whether upward or downward). There are various other classes of words with more than one stressed syllable; Chomsky has shown how the distinction between words like

(4) (i) ˌcompenˈsation (from to compensate), and
(ii) ˌconˌdenˈsation (from to condense).

can be predicted from the morphology. But equally well there is a very interesting distinction between the class of words, including BOTH (6 i) and (ii) which may be designated as having rising stress; as

(5) ˌobserˈvation, ˌorienˈtation, ˌsatyˈriasis, (ˌ)conˌfiguˈration, ˌpeneˈtralia, ˌineˈnarrable, ˌpapuˈliferous

--and those which we may call (following Jones, 1964 [1918]) double-stressed:

(6) 'syste'matic, 'repre'sent, 'peri'odically, 'bene'ficial, 'micro'scopic, 'un'moved, etc.

These words with two full stress-marks normally have accents assigned to both syllables. The so-called "stress alternation" is then produced by perfectly general rules of accent assignment and DELETION—in particular the "rhythm rule" (partially exemplified in Chomsky's alternating stress rule) which operates in the well known cases of

(7) 'un'known—ûnknòwn sóldier—the sôldier is ùnknówn
'home'made—hômemàde jám—the jâm is hòmemáde
'scouped 'up—sôuped ùp cár—the cár is sôuped úp

The stress distinctions between these two classes of words — those with more than one 'stressed' (LONG) syllable, one of which is SOLELY accented (normally) when the feature [+ACCENT] is assigned to the word, vs. those cacaucously accentable on two of the stressed syllables — has not (to my knowledge) been pointed out in quite this way before (& may not be confirmable); but cf. Jones, Pike, Bolinger.

The rules which assign accent (or emphasis) to some, but usually not all the full stressed syllables in a sense group depend on features of English structure which are usually manifested over text segments of greater than sentence length. It is typically the preceding context to which they are sensitive. The most interesting of these "context-sensitive rules" (if I may so extend that term) are those which shift the accent--or rather DELETE accents on certain items such that the next rightmost accent becomes nuclear. There is considerable nonsignificant variation in the assignment of accent to the eligible prenuclear syllables, which if taken into account can simplify our prosodic model. It used to be supposed, for example, that a sentence like (cf. Chomsky-Halle Sound Patterns)

(8)(i) Bîll's Brìtish hîstory tèacher is prétty F

was distinct from

(ii) Bill's British history teacher is pretty F

such that (i) meant Bill has a pretty teacher who teaches British History, and (ii) Bill has a pretty history teacher who is British. In fact the two deep structures collapse homophonously into the surface structure in (8)(i). Nine(ii) is just an optional stylistic variant where the rhythm rule has not applied.

Any syllebles after the nuclear-accented one whether stressed or not are unaccented and participate in the pitch "cadence"--cadence being a traditional term for what is sometimes called the nuclear fall of pitch. Sometimes the nuclear accent comes quite early and is followed by a long cadence, either in the same sense group or continued by a succession of enclitic adjunctions which are accentless. Examples of the latter type abound because medial and final vocatives and (cf. Bolinger) ascriptions are normally enclitic, and so are descriptive clauses subordinately conjoined to ascriptions, so we get

(9) "Good bye, Gilbert," she said, turning away and walking slowly but with finality down the long stair case.

Indeed a frequent solecism of journalistic prose is the inclusion in ascriptions of new information which ought to be accented but can't be because of the overriding effect of this rule: as

(10) The idea developed from experiments in mice, Dr. Mathé said in a preliminary report in the British Medical Journal.

---Harold Schmeck Jr.
The Semi-Artificial Man
(N.Y.: Walker, 1965, p. 179)

[Cadence traditionally referred to nuclear fall of pitch; i.e. the downward movement beginning on or just after the nuclear accent, so that we may designate [+CADENCE] by an "F" (for falling; it is more mnemonic and traditional. A fall-rise intonation has fall cadence with rising terminal endglide, (R). Lieberman's (1967) simplistic model utterly fails to distinguish between [-CAD, +ENDGLIDE] (i.e. rising intonation [R]), and fall-rise [+cadence, +ENDglide (FR). Further, since [+CAD, -ENDGLIDE] (F) corresponds to his "unmarked breath group," these examples of long cadences illustrate the irrelevance of his definition of "(-BG)" in terms of the fundamental frequency falling in "the last 150-200 msec of phonation."]

Examples of long cadence within a sense group are more apt to depend on "context sensitive" rules, and we must therefore look at enough of the context to see the operation of the rule. For example if we consider (11 ;) as an independent citation form we get

(11)(i) She used to have to find things and remind me of my appointments F*

Or, optionally, there can be a fall-rise pause after things, in which case the accent on find would of course become nuclear. Also, the accent on used is optionally deletable. Making both these changes we have

(ii) She used to have to find things FR and remind me of my appointments F*

I think you'll agree that the difference made by those stylistic options is small at the "level of delicacy" we are interested in. Contrast that with the very dramatic change in prosodic features imposed

* Where macrons instead of graves are used to show unaccented stressed syllables NOT in cadence.

by the operation of obligatory rules when the clause is restored to its context:

(12)(i) HIGGINS. [explaining about Eliza to his mother] ... she's useful. She knows where my things are, and remembers my appointments and so forth.

(ii) MRS. HIGGINS. How does your house keeper get on with her?

(iii) HIGGINS. Mrs. Pearce? Oh, she's jolly glad to get so much taken off her hands: for before Eliza came, she used to have to find things and remind me of my appointments.

—G.B. Shaw
Pygmalion Act III

The assignment of emphasis (roughly equal to Pike's pitch 1 or Trager and Smith's pitch 4) to the pronoun she and cadence to all the remainder of the sentence results from an oblicatory rule which says that used to have to find things and remind me of my appointments in (12)(iii) is to be de-accented as explicitly THEMATIC, being functionally anaphoric for or equivalent to knows where my things are, and remembers my appointments and so forth in (12)(i), while the two she's though morphemically identical are NOT (cf. Gleitman, 1961) discourse-equivalent because they are anaphoric for contrasting antecedents.

Of course to say this requires us to know that

(13) a) to find x = to know where x is,

b) to remind y of z = to remember z,

c) the contrast of tenses is non-rhematic (i.e. is not the "point" of the communication), and

d) Mrs. Pearce = $she_{iii} \neq she_{i}$ = Eliza

—which is to say that a very sophisticated level of syntactic and semantic analysis will be required in order for the correct assignment of prosodic features ever to be made by mechanical procedures.

Many of us heard a year ago about children's reactions concerning the anaphoric relations in sentences like

(14)(i) John X'ed Bill and he Y'd him (back).

We've begun playing around with sentences like this, without the adverb back. And since he and him are REGULARLY ambiguous, we use she/him, or he/her. Then if you have names in the forepart which happen to be ambiguous in terms of gender, you get

(ii) Leslie X'ed Gerry and then he Y'd her.

Or you can use various verb-like nonsense words as being more convincing then X and Y. Then asking which name is a male and which a female is an oblique way of getting at the antecedent identification. It appears that if you accent only the verb in the latter part: as

(15)(i) P$\hat{a}$t bl$\bar{u}$rged J$\hat{e}$an, and th$\bar{e}$n she skr$\acute{a}$ffed h$\grave{i}$m F

the order of feminine-masculine assigned to the names will be the same as on the pronouns, but if you emphasize the pronoun : as

(ii) D$\hat{o}$ll fl$\bar{o}\bar{o}$bed B$\hat{e}$rt and th$\bar{e}$n sh$\hat{e}$ cl$\bar{o}\bar{o}$fed h$\acute{i}$m F

the order is inverted just as surely as if you added back. Note that part of the gimmick is the use of semantically null verbs which MAY be in contrast (if the second is accented and thereby marked as "rheme") just as

(16)(i) John stabbed Bill and then he shot him F

or MAY be functionally synonymous---if unaccented and thereby marked as "theme" as in

(ii) John pushed Mary and then she shoved him F

We can make a conundrum by passivizing one verb: as

(17) #Pat blurged Gene and then he was skraffed by her F

where it is unclear who is who because its the nature of--pardon the expression--"functional sentence persepctive" that both he and by her when unaccented refer back to the subject noun. Even in real sentences we have to say

(18)(i) John hit Mary and then he was tripped by her F

(note accent on her) or

(ii) John hit Mary and then she was tripped by him F

(note accent on she)/ But enough of this violence. Let us turn to peaceful, loving examples like

(19) John regards Bill as incompetent FR and John castigates Bill as incompetent F

---a well formed and prosodically regular sentence, albeit marked by the lack of the normal pronominalization. Every word in the second part is obligatorily de-accented as identically equivalent with a functionally corresponding word in the first part except castigates, and the semantic range of this word permits it to contrast with regards, so it is emphasized. Then if

we consider

(20)(i) Bill's impression of John is that he is incompetent FR and John regards Bill as incompetent F

we see it is streightforward. Not only are the verbs equivalent, but John and Bill occur in the appropriate positions for contrastive emphasis. That is, even with a change of verb we can't say

(ii) *Bill's impression of John is that he is incompetent FR and John castigates Bill as incompetent F

This brings us at length to the example of interest: How do you read aloud the ~~xxx~~ following sentence:

(21)(i) John impresses Bill as incompetent, (ii) and John regards Bill as incompetent.

Quite clearly there is a correct answer--that is, the rules are not moot here--but it may require a bit of trial and error to get at it. This may be a case like multiple embedding where one has to use pencil and paper to apply one's native competence.

The problem is that the only possible meaning requires John and Bill in (ii) to contrast with Bill and John respectively in (i) but they are in the wrong positions to do so. Nevertheless, the one right reading of the line (allowing for minor stylistic options) is

(22)(i) John impresses Bill as incompetent FR (ii) and John regards Bill as incompetent F

This is a kind of literary style up with which most of us would never put. It's one of those grammatical but unacceptable sentences we heard about from my colleague, Terrence Moore.

That one was made up by conjoining a well known pair of illustrative sentences used by Chomsky, but it is not hard to find examples almost as egregious in real texts, as in

(23)(i) [One result of requiring excessively elaborate toxicity studies of a new drug] is unjustified delay in in obtaining a useful product if the drug proves to be valuable.

(ii) Many more lives may be lost by such delay FR than might be saved by excessive caution F

The problem here is that the parallel construction strongly suggest a double contrast (*lost by such delay: saved by excessive caution) but that is clearly wrong since excessive caution is here functionally synonymous with delay. Delay is, furthermore, de-accented because of the word delay in (23)(i). We can see from these examples that in applying the rules to assign correct prosodic features to sentences, logical and semantic relations are at least as centrally involved as phonological and syntactic ones.

Sometimes logical or syntactic inconsistences in the thought expressed makes the application of the prosodic rules ambiguous or moot; for such cases the notion "correct reading" is not well-defined. My favorite examples is from a label used by a west-coast bakery:

(24)(i) Extra Sour Dough Bread

which can only be read as

(ii) extra sourdough bread

--meaning a quantity of sourdough bread left over. There is no possible way of reading the phrase as it stands to make it mean what the bakery

intends; the bound morpheme sour in the compound noun 'sourdough cannot be separately modified by adverbial extra, and the only solution is textual, not prosodic--

(iii) êxtra-sôur sôurdôugh brêad

--which is hardly redolent with stylistic merit.

In this vein let me just suggest that the presence or absence of such prosodic solecisms may form an important part of the basis for judgements of literary quality. To that extent the widely accepted assumption that computers will always remain incapable of assessing literary merit may well be misconceived. On the other hand it must be conceded that once we go beyond straight factual prose and admit the whole corpus of fictive, dramatic, and poetical works, it becomes enormously difficult to separate the solecism of a hack from the coup of a genius.

I have tried to show that the apparently bewildering variations of accent placement in English are at least partially susceptible to prediction by "context sensitive" rules, and to suggest the motivation for and the nature of a few of the accentual features of a prosodic model in which stress, (postulated as lexical and defined in terms of (a) duration and (b) potential for accent,) is rigorously separated from the pitch phenomena: accent, emphasis and cadence.

This work was supported by the Office of Naval Research under contract N00014-67-A0111-0005 for Paralinguistic Studies of Speech.

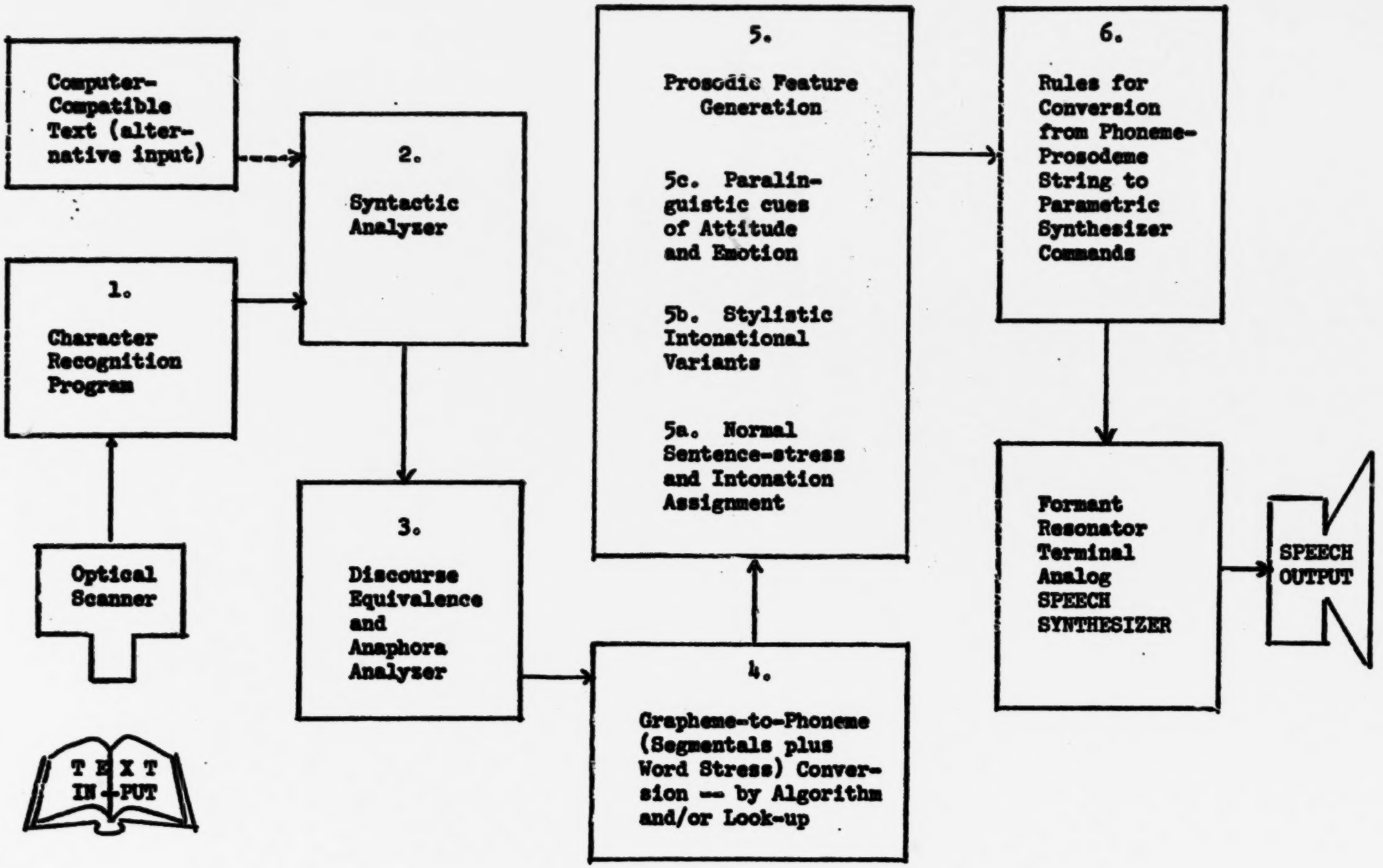

Figure 1. Schematic version of an automatic text-reading machine.

Ralph Vanderslice (UCLA), "Synthetic Elocution", RAND Linguistic Research Seminar, 23 February 1967.